CentER

Discussion paper

No. 2002-10

A GENERAL MODEL FOR REPEATED AUDIT CONTROLS USING MONOTONE SUBSAMPLING

By V.M. Raats, B.B. van der Genugten and J.J.A. Moors

February 2002

ISSN 0924-7815

A general model for repeated audit controls using monotone subsampling

Raats, V.M.[*], van der Genugten, B.B.* and J.J.A. Moors*

March 11, 2002

Abstract

In categorical repeated audit controls, fallible auditors classify sample elements in order to estimate the population fraction of elements in certain categories. To take possible misclassifications into account, subsequent checks are performed with a decreasing number of observations.

In this paper a model is presented for a general repeated audit control system, where k subsequent auditors classify elements into r categories. Two different subsampling procedures will be discussed, named 'stratified' and 'random' sampling. Although these two sampling methods lead to different probability distributions, it is shown, based on sufficiency, that the likelihood inferences are identical. The MLE and accompanying limit distributions are derived. The situations with undefined MLE are examined in detail; it is shown that an unbiased MLE can be acquired by stratified sampling.

Keywords: auditing, double checking, inspection errors, misclassifications, monotone missing data, repeated controls

Jel codes: C11, C13, C42, M41

[*]Tilburg University, P.O. Box 90153, 5000 LE Tilburg; email-adress: V.M.Raats@kub.nl

1 Introduction

Both the problem of missing data and the issue of misclassifications often occur in practice. Two main causes for missing observations are nonresponse and incomplete designs. While missing-by-design data is due to incomplete designs and therefore is intentionally created by the experimenter, this is usually not true for nonresponse. Misclassifications occur in quality control where a checking device has to classify objects in ($r \geq 2$) categories, e.g. 'good' or 'bad'. Sometimes it is known that the checking device is fallible, but it might be too expensive or just impossible to procure a better one. In many situations both problems occur simultaneously: not only some observations are missing, but there may be misclassifications as well. A practical example of missing-by-design data with possible misclassifications is a repeated audit control.

In a repeated audit control system one wants to draw conclusions about the fraction of elements in a population which belong to a certain category. In order to do this, an auditor classifies randomly sampled elements. However, misclassifications may occur, since the (usual) assumption that the auditor be infallible is dropped. To take these possible misclassifications into account, another fallible auditor checks a subsample of the already checked sample elements again. This procedure is repeated several times until the final k^{th} auditor, considered to be infallible, gives the true classification of some sample elements which already have been classified by all previous auditors. Conclusions about the population fractions have to be drawn based on the fallible and infallible audits. This kind of repeated audit control was introduced by Tenenbein (1970), who considered dichotomous data ($r = 2$) and two audit rounds ($k = 2$). Tenenbein (1972) extended the model to include categorical data ($r \geq 2$).

Our section 2 generalizes Tenenbein's work into a general control system for categorical data ($r \geq 2$) with monotone missing observations obtained from $k \geq 2$ audit rounds. Subsamples for subsequent auditors are obtained by using either 'stratified' or 'random' sampling. Though these different sampling methods lead to different probability distributions, it is shown in section 3 that the maximum likelihood estimators (MLE) for the main parameters are identical. However, only in case of 'stratified' sampling do these MLE appear to be unbiased. Section 4 discusses two different approaches to obtain confidence intervals for the population fractions based on the MLE, while section 5 considers the frequently occurring situations in which the MLE do not exist or are not unique. Section 6 contains an illustration, while the final section 7 contains the main conclusions.

2 A general model

2.1 Population model

Define the random variable I_0 as the true classification of a random sample element. The r possible classifications i_0 are denoted by $0, 1, ..., r-1$, while $p_{i_0} = \Pr(I_0 = i_0)$ denotes the population fraction of elements with true classification i_0.

A random element is classified by an auditor into one of the categories $\{0, 1, 2, ..., r-1\}$, leading to the random variable I_1. Hence a correct classification only occurs if $I_1 = I_0$. To take possible misclassifications into account, the same element may be categorized once more, now by another auditor. This procedure of subsampling is repeated, leading to classification I_j by auditor j, until the k^{th} auditor makes the final classification. Since this last auditor will be assumed to be an infallible expert, (s)he will always give the true classification: $I_k = I_0$.

The following notation will be used in the sequel to describe the different probabilities:

$$\begin{cases} p_{i_0 i_1 \ldots i_j} = & \Pr(I_0 = i_0, I_1 = i_1, ..., I_j = i_j) \quad j = 0, ..., k \\ \pi_{i_1 i_2 \ldots i_j} = & \Pr(I_1 = i_1, ..., I_j = i_j) \quad j = 1, ..., k. \end{cases}$$

It seems unrealistic to assume that classifications of subsequent auditors are independent, even if previous classifications are hidden: indeed, previous classifications reveal the difficulty of correctly classifying a given element. E.g. if many auditors judge an incorrect element to be correct, the error in the element probably is hard to detect. Hence we will need conditioning on previous classifications, to be denoted as follows:

$$\begin{cases} p_{i_j | i_0 i_1 \ldots i_{j-1}} = & \Pr(I_j = i_j | I_0 = i_0, ..., I_{j-1} = i_{j-1}) \quad j = 1, ..., k \\ \pi_{i_j | i_1 \ldots i_{j-1}} = & \Pr(I_j = i_j | I_1 = i_1, ..., I_{j-1} = i_{j-1}) \quad j = 2, ..., k. \end{cases}$$

Since the last auditor is infallible ($I_k = I_0$), it follows $\pi_{i_1 i_2 \ldots i_k} = p_{i_0 i_1 \ldots i_k} = p_{i_0 i_1 \ldots i_{k-1}}$. There exist the following relations between the two sets of parameters:

$$\begin{cases} \text{(a)} \quad \pi_{i_1 i_2 \ldots i_k} = & p_{i_0} \cdot p_{i_1 | i_0} \cdot p_{i_2 | i_0 i_1} \cdot \ldots \cdot p_{i_{k-1} | i_0 i_1 \ldots i_{k-2}} \\ \text{(b)} \quad \pi_{i_1 i_2 \ldots i_k} = & \pi_{i_1} \cdot \pi_{i_2 | i_1} \cdot \pi_{i_3 | i_1 i_2} \cdot \ldots \cdot \pi_{i_k | i_1 \ldots i_{k-1}} \\ \text{(c)} \quad p_{i_0} = p_{i_k} = & \sum_{i_1 .. i_{k-1}} \pi_{i_1 i_2 .. i_k} \quad . \end{cases} \tag{1}$$

Finally the following shorthand notations are introduced:

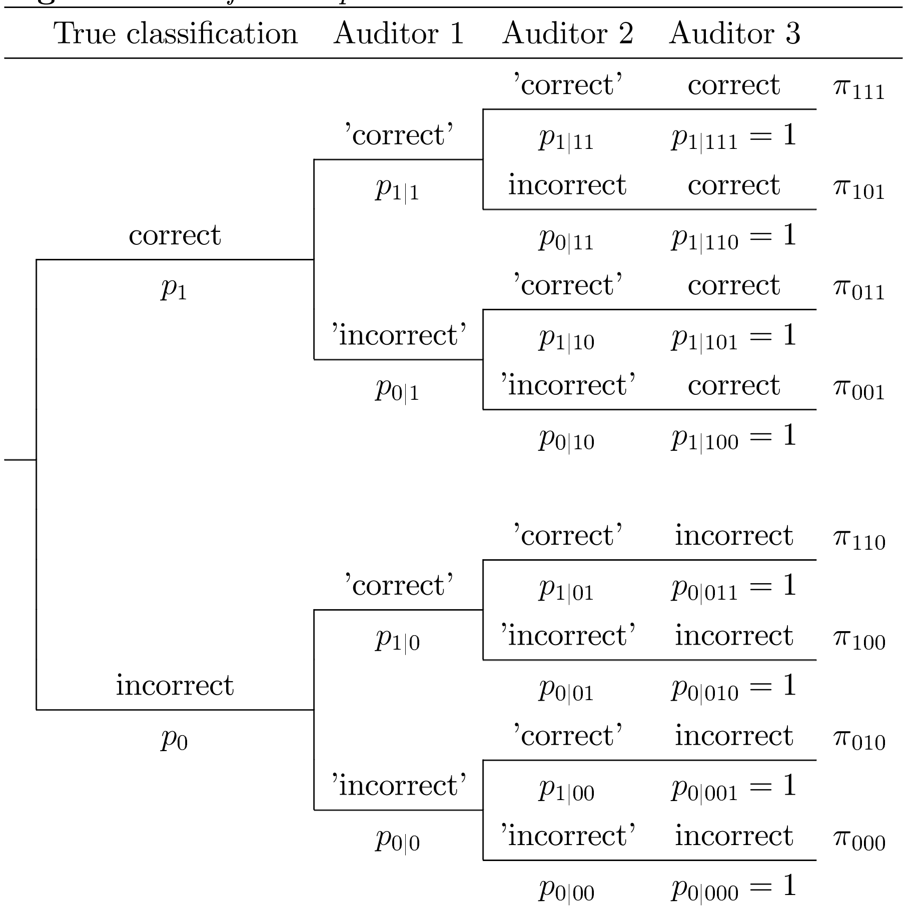
$$\left\{\begin{array}{ll} c: & \text{one of the } r^{j-1} \text{ possible classifications } i_1 i_2 ... i_{j-1} \text{ by the first } j-1 \text{ auditors} \\ \pi_c^{(j)}: & \text{row vector of } r \text{ probabilities } \pi_{ci_j}\ (i_j = 0, 1, ..., r-1) \\ \pi^{(j)}: & (r^{j-1} \times r) \text{ matrix with rows } \pi_c^{(j)} \\ \pi_c^{(j|j-1)}: & \text{row vector of } r \text{ probabilities } \pi_{i_j|c}\ (i_j = 0, 1, ..., r-1) \\ \pi^{(j|j-1)}: & (r^{j-1} \times r) \text{ matrix with rows } \pi_c^{(j|j-1)} \\ p_{i_0 c}^{(j|j-1)}: & \text{row vector of } r \text{ probabilities } p_{i_j|i_0 c}\ (i_j = 0, 1, ..., r-1) \\ p^{(j|j-1)}: & (r^j \times r) \text{ matrix with rows } p_{i_0 c}^{(j|j-1)} \quad . \end{array}\right.$$

The matrices are constructed with columnwise and rowwise decreasing classifications. These notations are illustrated below for the simplest example without lack of generality.

Illustration ($r = 2$ and $k = 3$)

Consider a population which consists of incorrect ($i_0 = 0$) and correct elements ($i_0 = 1$). In order to draw conclusions about the population fraction of incorrect elements, a repeated audit control with three rounds is performed of which the last is infallible. Figure 1 gives an overview of all relevant probabilities; see also Figure 2.

Figure 1 *Classification probabilities*



$$\pi^{(1)} = \begin{pmatrix} \pi_1 & \pi_0 \end{pmatrix},$$

$$\pi^{(2)} = \begin{pmatrix} \pi_1^{(2)} \\ \pi_0^{(2)} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{10} \\ \pi_{01} & \pi_{00} \end{pmatrix}, \qquad \pi^{(2|1)} = \begin{pmatrix} \pi_1^{(2|1)} \\ \pi_0^{(2|1)} \end{pmatrix} = \begin{pmatrix} \pi_{1|1} & \pi_{0|1} \\ \pi_{1|0} & \pi_{0|0} \end{pmatrix}$$

$$\pi^{(3)} = \begin{pmatrix} \pi_{11}^{(3)} \\ \pi_{10}^{(3)} \\ \pi_{01}^{(3)} \\ \pi_{00}^{(3)} \end{pmatrix} = \begin{pmatrix} \pi_{111} & \pi_{110} \\ \pi_{101} & \pi_{100} \\ \pi_{011} & \pi_{010} \\ \pi_{001} & \pi_{000} \end{pmatrix}, \qquad \pi^{(3|2)} = \begin{pmatrix} \pi_{11}^{(3|2)} \\ \pi_{10}^{(3|2)} \\ \pi_{01}^{(3|2)} \\ \pi_{00}^{(3|2)} \end{pmatrix} = \begin{pmatrix} \pi_{1|11} & \pi_{0|11} \\ \pi_{1|10} & \pi_{0|10} \\ \pi_{1|01} & \pi_{0|01} \\ \pi_{1|00} & \pi_{0|00} \end{pmatrix}$$

$$p^{(2|1)} = \begin{pmatrix} p_{11}^{(2|1)} \\ p_{10}^{(2|1)} \\ p_{01}^{(2|1)} \\ p_{00}^{(2|1)} \end{pmatrix} = \begin{pmatrix} p_{1|11} & p_{0|11} \\ p_{1|10} & p_{0|10} \\ p_{1|01} & p_{0|01} \\ p_{1|00} & p_{0|00} \end{pmatrix}, \quad p^{(1|0)} = \begin{pmatrix} p_1^{(1|0)} \\ p_0^{(1|0)} \end{pmatrix} = \begin{pmatrix} p_{1|1} & p_{0|1} \\ p_{1|0} & p_{0|0} \end{pmatrix}$$

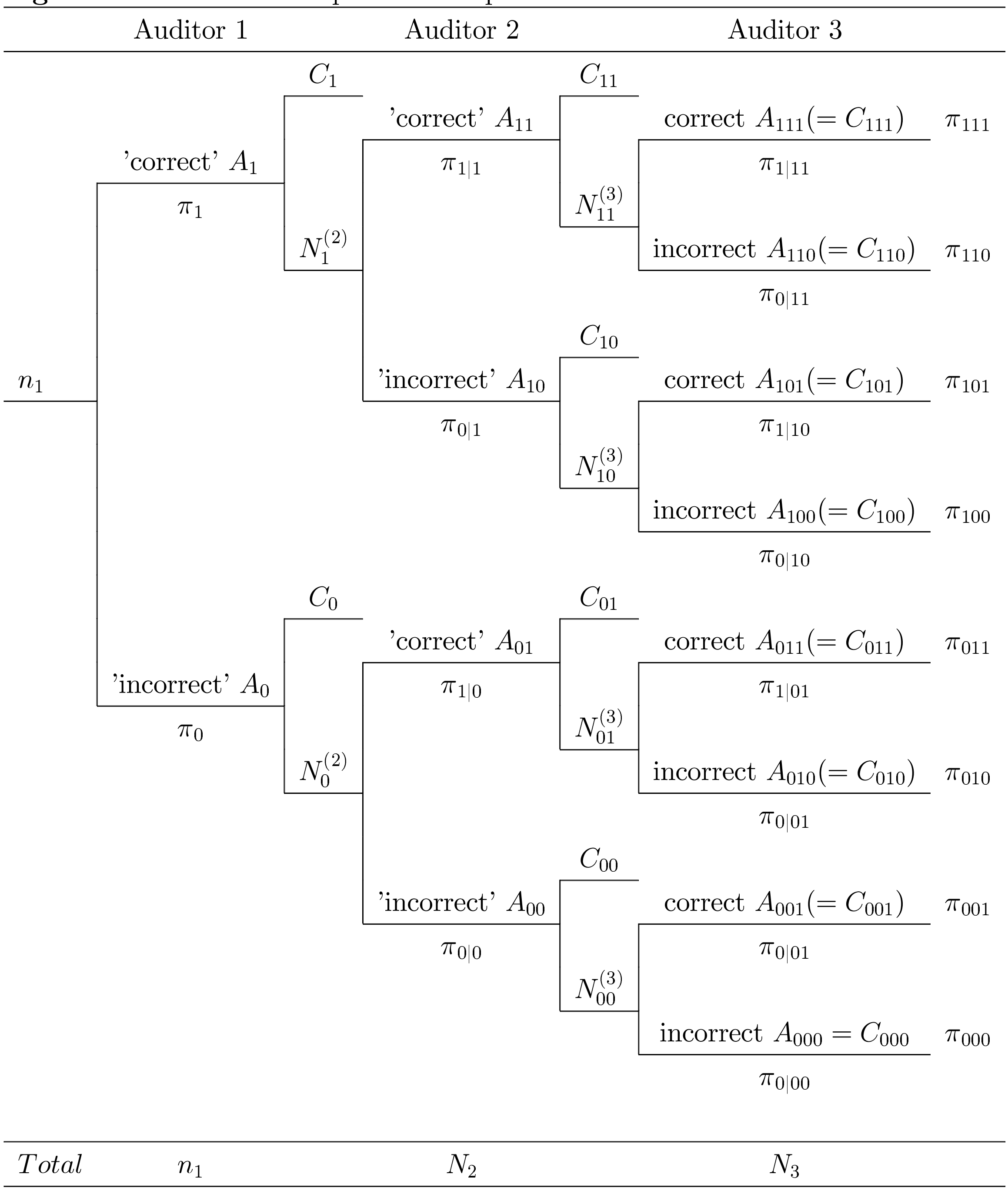
2.2 Sample information

Auditor 1 classifies the elements of a random sample (drawn with replacement) of predetermined size n_1; a subsample of (possibly random) size $N_2 \leq n_1$ is checked again by auditor 2, and so on: auditor j checks $N_j \leq N_{j-1}$ elements ($j = 3, ..., k$). Hence, N_k elements are classified by all auditors, $N_j - N_{j+1}$ elements by precisely the first j auditors. Such a pattern of observations is called a monotone missing data pattern; see Little and Rubin (1987). Note that here missing-by-design occurs.

Let A_c denote the number of elements classified by the first $j-1$ auditors as $c = i_1 ... i_{j-1}$. Of these, $N_c^{(j)} \leq A_c$ are observed by auditor j; the remainder $C_c = A_c - N_c^{(j)}$ is not further investigated. The classification frequencies of auditor j are A_{ci_j} to be combined into the vector $A_c^{(j)}$. These r^{j-1} vectors can be collected into the matrix $A^{(j)}$, presenting all frequencies, observed by the first j auditors. These notations agree with the notations for the parameters π. The k matrices $A^{(j)}$ summarize the complete sample information.

Illustration ($r = 2$ and $k = 3$) continued.

Figure 2 Classification frequencies and probabilities



2.3 Sampling methods

An important aspect of a repeated audit control is the way in which it is decided which sample elements have to be checked again. In general, we allow the sample sizes to depend on the preceding results. Two different sampling methods will be discussed here: stratified and random sampling. In case of stratified sampling, the sample size

$N_c^{(j)}$ in round j from any given classification c is determined separately, while in random sampling only the total N_j over all these r^{j-1} classifications is prescribed. More precisely, let $\mathcal{A}^{(j)}$ denote the outcome space of $A^{(j)}$, while $f_c^{(j)}$ and g_j are given functions from $\mathcal{A}^{(1)} \cdot \mathcal{A}^{(2)} \cdot \ldots \cdot \mathcal{A}^{(j-1)}$ into respectively $\mathbb{N} \cup \{0\}$ and $\mathbb{N}$ for all c and j. Then the two methods can be described as follows:

$$\begin{cases} \text{stratified sampling:} & N_c^{(j)} = f_c^{(j)}(A^{(1)}, \ldots, A^{(j-1)}) \\ \text{random sampling:} & N_j = g_j(A^{(1)}, \ldots, A^{(j-1)}) \quad . \end{cases}$$

Hence as soon as $A^{(j-1)}$ is known, the $N_c^{(j)}$ and N_j are given. Of course, the realization of the total sample size in round j also has to be positive for stratified sampling: $N_j = \sum\limits_c N_c^{(j)} > 0$.

In most cases sample sizes will only depend on the previous round frequencies, so that $N_j = g_j(A^{(j-1)})$, e.g.; the simplest situation occurs when all the sample sizes are fixed predetermined numbers. This is the sampling method which is usually assumed in the existing literature on repeated audit controls.

3 Distributions and MLE

3.1 Stratified sampling

All the following results are derived under the assumption of sampling with replacement. The convention that the multinomial distribution $M(0;.)$ is concentrated in 0 will be adopted.

Theorem 1.

In case of stratified sampling the joint sample distribution is characterized by the following multinomial distributions:

$$\begin{cases} \mathcal{L}(A^{(1)}) = M(n_1; \pi^{(1)}), \\ \mathcal{L}(A_c^{(j)} | N_c^{(j)} = n_c^{(j)}) = M(n_c^{(j)}; \pi_c^{(j|j-1)}), \text{ for all } r^{j-1} \text{ possible } c, \quad j = 2, \ldots, k \quad . \end{cases} \tag{2}$$

and the likelihood function $L(\pi^{(1)}, \pi_1^{(2|1)}, \ldots, \pi_c^{(k|k-1)}; a^{(1)}, .., a^{(k)})$ is obtained by multiplying all probabilies corresponding with the $(1 - r^k)/(1 - r)$ multinomials in (2).

Proof:

Equation (2) is obvious. Further, because the f_c^j are given functions,

$$\mathcal{L}(A_c^j | A^{(1)}, \ldots, A^{(j-1)}) = \mathcal{L}(A_c^{(j)} | N_c^{(j)})$$

holds for all c and j, while these distributions are conditionally independent for different c. This implies the second statement. $\square$

The corresponding log-likelihood follows at once:

$$\log\, L(\pi^{(1)}, \pi_1^{(2|1)}, ..., \pi_c^{(k|k-1)}; a^{(1)}, .., a^{(k)}) = \sum_{i_1} a_{i_1} \log \pi_{i_1} + \sum_{j=2}^{k} \sum_{ci_j} a_{ci_j} \log \pi_{i_j|c} \qquad (3)$$

as well as the MLE for all parameters involved:

$$\begin{cases} \widehat{\Pi}^{(1)} & = \dfrac{A^{(1)}}{n_1}, \\ \widehat{\Pi}_c^{(j|j-1)} & = \dfrac{A_c^{(j)}}{N_c^{(j)}}, \quad \text{for all } r^{j-1} \text{ possible } c, \quad j = 2, ..., k \quad . \end{cases} \qquad (4)$$

These MLE are the usual multinomial estimators (see Little and Rubin (1987) for more details). Since the parameters of interest p_{i_0} are functions of $(\pi_{i_1}, \pi_{i_j|c})$ (see (1)), the MLE for p_{i_0} are functions of the derived MLE (4):

$$\widehat{P}_{i_0} = \widehat{P}_{i_k} = \sum_{i_1..i_{k-1}} \widehat{\Pi}_{i_1 i_2..i_k} = \sum_{i_1..i_{k-1}} \widehat{\Pi}_{i_1} \cdot \widehat{\Pi}_{i_2|i_1} \cdot ... \cdot \widehat{\Pi}_{i_k|i_1..i_{k-1}} \quad . \qquad (5)$$

However, the MLE for the conditional classification probabilities $\pi_{i_j|c}$ are not defined when $N_c^{(j)} = 0$. This is asymptotically irrelevant but highly relevant in practice! Although the probability of an unidentified estimator tends asymptotically to zero, practical repeated audit controls usually have small final sample sizes due to the high costs of the last auditor. Undefined MLE are (in general) frequently occurring and it is important to have a good estimation procedure which can handle these situations. Section 5 examines possible procedures for undefined MLE more closely.

Note that the auditors' error probabilities can be derived from (1), (4) and (5) as well; e.g.

$$\widehat{P}_{i_1|i_0} = \widehat{P}_{i_1|i_k} = \frac{\widehat{P}_{i_1 i_k}}{\widehat{P}_{i_k}} = \frac{\sum_{i_2..i_{k-1}} \widehat{\Pi}_{i_1 i_2..i_k}}{\sum_{i_1..i_{k-1}} \widehat{\Pi}_{i_1 i_2..i_k}} = \frac{\sum_{i_2..i_{k-1}} \widehat{\Pi}_{i_1} \cdot \widehat{\Pi}_{i_2|1} \cdot ... \cdot \widehat{\Pi}_{i_k|i_1..i_{k-1}}}{\sum_{i_1..i_{k-1}} \widehat{\Pi}_{i_1} \cdot \widehat{\Pi}_{i_2|1} \cdot ... \cdot \widehat{\Pi}_{i_k|i_1..i_{k-1}}} \quad .$$

3.2 Random sampling

Although the $N_c^{(j)}$ are deterministic conditionally on the previous classifications in the case of stratified sampling, this is not true for random sampling and the characteristic distributions differ for the two sampling methods. Let $N^{(j)}$ denote the vector of all r^{j-1} scalars $N_c^{(j)}$.

Theorem 2.
In case of random sampling the joint sample distribution is characterized by the following multinomial distributions:

$$\begin{cases} \mathcal{L}(A^{(1)}) = M(n_1; \pi^{(1)}), \\ \mathcal{L}(N^{(j)}|A^{(j-1)} = a^{(j-1)}, N_j = n_j) = M(n_j; \dfrac{vec(a^{(j-1)})}{n_{j-1}}), & j = 2, ..., k \\ \mathcal{L}(A_c^{(j)}|N_c^{(j)} = n_c^{(j)}) = M(n_c^{(j)}; \pi_c^{(j|j-1)}), \quad \text{for all } r^{j-1} \text{ possible } c, & j = 2, ..., k \end{cases} \quad . \tag{6}$$

but the likelihood inferences for both random and stratified sampling are identical.
Proof:
The conditional multinomial distribution functions (6) are again straightforward. The likelihood is now acquired by multiplying all the $(1-r^k)/(1-r)+k-1$ conditionally independent multinomial distributions:

$$\begin{aligned} &L(\pi^{(1)}, \pi_1^{(2|1)}, ..., \pi_c^{(k|k-1)}; a^{(1)}, .., a^{(k)}, n^{(2)}, ..., n^{(k)}) \\ &\quad = L(A^{(1)})L(N^{(2)}|A^{(1)}, n_2)L(A_0^{(2)}|N_0^{(2)})L(A_1^{(2)}|N_1^{(2)}) \cdot ... \cdot L(A_c^{(k)}|N_c^{(k)}) \quad . \end{aligned}$$

The conditional distribution functions for the classification quantities $A^{(1)}$ and $A_c^{(j)}$ are identical for random and stratified subsampling. Therefore the likelihood functions of the two sampling methods differ only by the additional conditional distribution functions of the sample sizes $N^{(j)}$ in case of random sampling. Since these distribution functions do not involve any parameters of interest, the distributions of the $N^{(j)}$ can be ignored for likelihood inferences about the parameters: $A^{(1)}$ and $A_c^{(j)}$ are sufficient for $\pi_{i_1}$and $\pi_{i_j|c}$, respectively. $\square$

Although, we have mainly focussed on point estimators so far, in auditing practice a lot of decisions are based on confidence intervals for the p_{i_0} rather than on the point estimates.

4 Confidence intervals

4.1 Finite samples

For a standard audit with an infallible auditor ($k=1$) and dichotomous data ($r=2$) the upper $(1-\alpha)-$confidence limit for p_{i_0}, denoted by $p_{i_0}^u$ is the regular binomial confidence limit

$$p_{i_0}^u = \underset{p_{i_0}}{max} \left\{ p_{i_0} : \Pr(\widehat{P}_{i_0} \leq \widehat{p}_{i_0}|p_{i_0}) \geq \alpha \right\} \quad . \tag{7}$$

The generalization for $r = 2$ and $k = 2$ is simply:

$$p_0^u = \underset{p_{i_0}}{max} \left\{ p_0, p_{1|0}, p_{0|1} : \Pr(\widehat{P}_0 \leq \widehat{p}_0 | p_0, p_{1|0}, p_{1|1}) \geq \alpha \right\} \quad . \tag{8}$$

To determine this upper limit, the maximum $p_0^u|p_{1|0}, p_{0|1}$ of (8) for fixed $p_{1|0}$ and $p_{0|1}$ has to be calculated for all possible values of the nuisance parameters. Subsequently, p_0^u is determined as the maximum of all $p_0^u|p_{1|0}, p_{0|1}$. Compare Moors et al. (2000).
It is straightforward to generalize (8) for $r \geq 2$ and $k \geq 2$:

$$p_{i_0}^u = \underset{p_{i_0}}{max} \left\{ p_{i_0}, p^{(j|j-1)} : \Pr(\widehat{P}_{i_0} \leq \widehat{p}_{i_0} | p_{i_0}, p^{(j|j-1)}, \ j = 1, ..., k-1) \geq \alpha \right\} \quad .$$

The determination of $p_{i_0}^u$ goes in a similar way as for the case where $r = 2$ and $k = 2$. A disadvantage of this method is the worst case approach: when determining the upper limit all situations (i.e. all values of the nuisance parameters) are considered and the most unfavorable one is taken as a starting point. All possible situations also include the situation in which each fallible auditor deliberately classifies all elements in the same category regardless of the true and previous classifications: i.e. for $j = 1, ..., k-1$ the elements of $p^{(j|j-1)}$ consist solely of zeros and ones. As a consequence all elements will be classified in exactly the same way by the first $k-1$ auditors: $i_1^*, ..., i_{k-1}^*$. In this case the MLE for p_{i_0} (5) reduces to

$$\widehat{P}_{i_0} = \quad \widehat{P}_{i_k} = \widehat{\Pi}_{i_1^* i_2^* .. i_{k-1}^* i_k} = \quad \widehat{\Pi}_{i_k | i_1^* .. i_{k-1}^*} = \frac{A_{i_1^* ... i_{k-1}^* i_k}}{N_k} \quad .$$

The latter is just the estimator in case of an ordinary audit with only an infallible auditor who checks n_k elements. So $p_{i_0}^u|p^{(j|j-1)}$ is solely based on the classifications by the last infallible auditor and the fallible classifications are disregarded completely. Therefore it coincides with the upper limit of a standard audit by an infallible auditor who checks n_k elements (7). As a consequence $p_{i_0}^u$, which is the maximum of all $p_{i_0}^u|p^{(j|j-1)}$ will be at least as high as (7) and the repeated audit control is in this sense useless: the fallible classifications cost money but do not provide more accurate estimates.

So although the described method enables us to find confidence limits for **finite** samples, these confidence limits will be very high since the -often unlikely- worst case is taken as a starting point. This conclusion is in line with the results of Moors et al. (2000) and Raats and Moors (2000).

4.2 Limit distribution

A widely applied approach to construct confidence intervals is based on the limit distribution of the MLE.

Theorem 3.
Under the assumption $\frac{N_c^{(j)}}{n_1} \xrightarrow{p} b_c^{(j)}$ with $b_c^{(j)}$ a constant depending on c,

$$\sqrt{n_1}(\widehat{P}_{i_0} - p_{i_0}) \xrightarrow{\mathcal{L}} N(0, \sigma_{i_0}^2) \tag{10}$$

with

$$\sigma_{i_0}^2 = \sigma_{i_k}^2 = \sum_{i_1 \ldots i_{k-1}} Var(\widehat{\Pi}_{i_1 \ldots i_{k-1} i_k}) + \sum_{i_1 \ldots i_{k-1} \neq i'_1 \ldots i'_{k-1}} Cov(\widehat{\Pi}_{i_1 \ldots i_{k-1} i_k}, \widehat{\Pi}_{i'_1 \ldots i'_{k-1} i_k})$$

define w =min$\{j : i_j \neq i'_j$ for $i_1 \ldots i_k$ and $i'_1 \ldots i'_k\}$ then

$$\text{Cov}(\widehat{\Pi}_{i_1 \ldots i_k}, \widehat{\Pi}_{i'_1 \ldots i'_k}) = \begin{cases} -\pi_{i_1 \ldots i_k} \pi_{i'_1 \ldots i'_k} & \text{if } w = 1 \\ \frac{\pi_{i_1 \ldots i_k}}{\pi_{i_1}} \frac{\pi_{i'_1 \ldots i'_k}}{\pi_{i_1}} \pi_{i_1}(1 - \pi_{i_1}) + \\ \sum_{j=2}^{w-1} \frac{\pi_{i_1 \ldots i_k}}{\pi_{i_j|c}} \frac{\pi_{i'_1 \ldots i'_k}}{\pi_{i_j|c}} \frac{\pi_{i_j|c}(1 - \pi_{i_j|c})}{b_c^{(j)}} - \frac{\pi_{i_1 \ldots i_k} \pi_{i'_1 \ldots i'_k}}{b_{i_1 \ldots i_{w-1}}^{(w)}} & \text{if } 1 < w \leq k \\ \left(\frac{\pi_{i_1 \ldots i_k}}{\pi_{i_1}}\right)^2 \pi_{i_1}(1 - \pi_{i_1}) + \sum_{j=2}^{k} \left(\frac{\pi_{i_1 \ldots i_k}}{\pi_{i_j|c}}\right)^2 \frac{\pi_{i_j|c}(1 - \pi_{i_j|c})}{b_c^{(j)}} & \text{else} \end{cases}$$

Proof: see appendix A. □

Now the standard techniques can be applied to construct confidence intervals. Tenenbein (1970,1971,1972) used the variance of the limit distribution $\sigma_{i_0}^2$ as a measure of accuracy of the repeated audit control. However, as mentioned before, asymptotics are often not relevant for these types of controls.

Neither of the two methods for constructing confidence intervals which are discussed so far, appears to be very useful. Therefore other approaches should be considered such as a decision theoretic or Bayesian approach. Raats & Moors (2000) analyzed the latter for $r = 2$ and $k = 2$. They specified beta distributions as priors for the p_{i_0} and $p_{i_1|i_0}$. Though this approach seems to give acceptable results, the complexity of the required calculations increases very fast with the number of categories r and the number of audit rounds k. Schafer (1997) discusses a Bayesian approach for multinomial monotone missing data. He proposes Dirichlet distributions as priors for the Π_{i_1} and $\Pi_{i_j|c}$ to obtain posteriors for these parameters. Technically it is also possible to do this for our models though it seems more logical to formulate priors for P_{i_0} (and the $P_{i_j|i_0c}$) rather than for Π_{i_1} and $\Pi_{i_j|c}$ since the parameters of interest are p_{i_0}.

5 Undefined MLE

Though the MLE have nice asymptotic properties and are logically interpretable, a major drawback is that they will be frequently undefined in practice (depending on the sampling method). The MLE for the population fraction are undefined when auditor j does not classify at least one sample element of each previously occurring classification pattern, i.e. $n_c^{(j)} = 0$ while $a_c > 0$. The situation $n_c^{(j)} = 0$ can be divided into structural zeros and unstructural zeros (see Bishop et al. (1975)). Unstructural zeros are caused by chance while structural zeros are caused by a priori model restrictions such as $\pi_c = 0$. In this paper we extend this last definition to include the situation $n_c^{(j)} = 0$ when $a_c > 0$ where the elements with previous classification c are intentionally excluded from the j^{th} sample ($N_c^{(j)} = f_c^{(j)}(A^{(1)}, .., A^{(j-1)}) = 0$) because another check would not provide additional information.

Consider for example a population which consists of correct ($i_0 = 1$) and incorrect elements ($i_0 = 0$). A repeated audit control takes place with only one fallible auditor ($k = 2$). The fallible auditor is a priori known never to misclassify correct elements ($p_{1|1} = 1$) but (s)he might make mistakes with incorrect elements. As a consequence an element which the first auditor classifies as incorrect is per definition incorrect. An additional check of such an element does not provide extra information and is therefor useless. A logical choice is $N_0^2 = 0$. Though $\widehat{\Pi}_{1|0}$ is now undefined according to (4), this is not a problem since it is a priori known that $\pi_{1|0} = 0$.

In general, structural zeros don't cause problems because they are caused themselves by model assumptions about the parameters. Unstructural zeros, however, are the cause of some problems. Fortunately, unstructural zeros can be avoided completely by using a specific kind of stratified sampling: stratified sampling with $N_c^{(j)} > 0$ when $a_c > 0$. In these cases the MLE for p_{i_0} are always uniquely defined and are even unbiased.

Theorem 4.
$E\{\widehat{P}_{i_0}\} = p_{i_0}$ if $N_c^{(j)} > 0$ when $a_c > 0$
Proof:
Under this kind of stratified sampling the MLE $\widehat{\Pi}_{i_j|c}(4)$ can still be undefined. However, the preceding factor in the last term of (5) $\widehat{\Pi}_{i_{j-1}|i_1 \ldots i_{j-2}}$ is per definition 0 when $N_c^{(j)} = 0$. As a consequence the corresponding term $\widehat{\Pi}_{i_1 \ldots i_k}$ of $\widehat{P}_{i_0}(5)$ is zero. So the MLE $\widehat{P}_{i_0}$ is

not undefined, even in case of undefined conditional classification probabilities. From the relations

$$\begin{aligned} E\{\widehat{\Pi}_{i_1 i_2..i_j}\} &= E\{\widehat{\Pi}_c \cdot \widehat{\Pi}_{i_j|c}\} && = E\{\widehat{\Pi}_c\} E\{\frac{A_{ci_j}}{N_c^{(j)}} | N_c^{(j)}\} \\ &= E\{\widehat{\Pi}_c\} \cdot \pi_{i_j|c} && = E\{\widehat{\Pi}_{i_1 \ldots i_{j-1}}\} \cdot \pi_{i_j|i_1 \ldots i_{j-1}} \end{aligned}$$

it follows by repeated application that

$$E\{\widehat{\Pi}_{i_1 i_2..i_j}\} = \pi_{i_1 i_2..i_j}$$

Now the unbiasedness of $\widehat{P}_{i_0}$ follows straightforward from definition (5):

$$E\{\widehat{P}_{i_0}\} = \sum_{i_1..i_{k-1}} E\{\widehat{\Pi}_{i_1 i_2..i_k}\} = \sum_{i_1..i_{k-1}} \pi_{i_1 i_2..i_k} = p_{i_0} \qquad \square.$$

A disadvantage of this kind of stratified sampling is that the required final sample size can be quite large since the last sample has to include at least one element of all previous realized classifications. This could be an argument to apply a different sampling method which could still lead to unstructural zeros. The following section shows that the procedure for handling situations with undefined MLE is indeed important.

6 Illustration *(k=2, r=2)*

A population consists of correct ($i_0 = 1$) and incorrect ($i_0 = 0$) elements. In order to estimate p_0, a repeated audit control is performed by two auditors. Random sampling is applied with n_2 being a fixed number: $N_2(a_1, a_0) = n_2$. There are no prior assumptions about the quality of the first auditor, i.e. about the misclassification probabilities. The characteristic sample distributions (6) are reduced to:

$$\begin{cases} \mathcal{L}(A_1, A_0) = M(n_1; \pi_1, \pi_0), \\ \mathcal{L}(N^{(2)}|A_1 = a_1, A_0 = a_0) = M(n_2; \frac{a_1}{n_1}, \frac{a_0}{n_1}), \\ \mathcal{L}(A_{11}, A_{10}|N_1^{(2)} = n_1^{(2)}) = M(n_1^{(2)}; \pi_{1|1}, \pi_{0|1}), \\ \mathcal{L}(A_{01}, A_{00}|N_0^{(2)} = n_0^{(2)}) = M(n_0^{(2)}; \pi_{1|0}, \pi_{0|0}). \end{cases}$$

and the MLE (5) follows.

Both Tenenbein (1970) and Moors (2000) derived this MLE. Tenenbein (1970) noted that the MLE for p_0 is undefined when either $N_0^{(2)}$ or $N_1^{(2)}$ equals 0, but he concluded that the probability of this occurring is quite small unless n_2 is small and π_1 or π_0 is close to zero. However, these cases are of importance for calculating upper confidence limits. Moors (1999) derived the MLE independently from Tenenbein and paid special attention to the cases of undefined MLE. To determine the MLE in the case of unstructural zeros, he made the extra assumption $p_{1|0} = 1 - p_{0|1}$. This resulted in the following estimator:

$$\widehat{P}_0 = \begin{cases} \dfrac{A_{10}}{N_1^{(2)}} & \text{if } N_0^{(2)} = 0 \\ \dfrac{A_0}{n_1}\dfrac{A_{00}}{N_0^{(2)}} + \dfrac{A_1}{n_1}\dfrac{A_{10}}{N_1^{(2)}} & \text{if } 0 < N_0^{(2)} < n_2 \\ \dfrac{A_{00}}{N_0^{(2)}} & \text{if } N_0^{(2)} = n_2 \end{cases} \tag{11}$$

The main expression consists of two terms which have a logical interpretation. The first term is the fraction of elements which are classified as 'incorrect' by the first auditor times the estimated probability that they are actually incorrect. The second term is the fraction of elements which are classified as 'correct' by the first auditor times the estimated probability that they are actually incorrect. If unstructural zeros occur, all information of the fallible auditor is discarded.

The major disadvantage of Moors' estimator $\widehat{P}_0$ is that it does not coincide with the MLE for the reduced models. In a reduced model, one misclassification probability, either $p_{1|0}$ or $p_{0|1}$, is a priori set to zero. It can be shown that Moors' estimator does not coincide with the MLE of the two reduced models if unstructural zeros occur. Therefore, a slightly modified estimator is proposed:

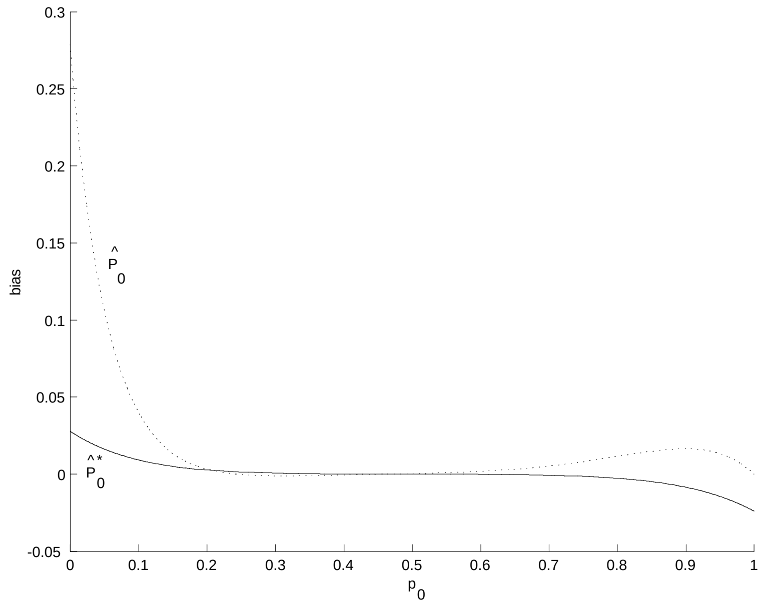
$$\widehat{P}_0^* = \begin{cases} \dfrac{A_0}{n_1} + \dfrac{A_1}{n_1}\dfrac{A_{10}}{N_1^{(2)}} & \text{if } N_0^{(2)} = 0 \\ \dfrac{A_0}{n_1}\dfrac{A_{00}}{N_0^{(2)}} + \dfrac{A_1}{n_1}\dfrac{A_{10}}{N_1^{(2)}} & \text{if } 0 < N_0^{(2)} < n_2 \\ \dfrac{A_0}{n_1}\dfrac{A_{00}}{N_0^{(2)}} & \text{if } N_0^{(2)} = n_2 \end{cases} \tag{12}$$

This is the only estimator which coincides with the MLE of the reduced models. In order to see whether the differences between (11) and (12) are relevant, a comparison is made based on the bias. By taking conditional expectations (see appendix B) it follows:

$$\begin{aligned} \text{Bias}(\widehat{P}_0) &= (1 - \frac{n_2}{n_1})(\pi_1^{n_2}(\pi_{0|1} - \pi_{00} - \pi_{10}) + \pi_0^{n_2}(\pi_{0|0} - \pi_{00} - \pi_{10})) \\ \text{Bias}(\widehat{P}_0^*) &= (1 - \frac{n_2}{n_1})(\pi_1^{n_2}\pi_{01} - \pi_0^{n_2}\pi_{10}) \end{aligned}$$

The bias of both estimators depends on the classification probabilities and the sample sizes. The bias is reduced by increasing n_2 or decreasing n_2/n_1. This means that the bias is smaller if more infallible information is acquired or if the fraction of fallible information decreases. The bias of $\widehat{P}_0^*$ decreases when the first auditor is more accurate; it is even unbiased in the case of an infallible first auditor. The latter is not true for estimator (11). Figure 3 shows that the difference between the estimators can be quite substantial.

Figure 3 Bias of the estimators $\widehat{P}_0$ and $\widehat{P}_0^*$ as function of p_0



This graph shows the bias of estimators (11) and (12) for $n_1 = 50$, $n_2 = 10$, $p_{1|0} = 0.05$ and $p_{0|1} = 0.10$. In particular for low values of p_0, use of the modified estimator $\widehat{P}_0^*$ reduces the bias considerably.

7 Conclusions

A general framework for repeated audit controls was introduced for categorical data with $r \geq 2$ levels. Monotone sampling (cf. Little and Rubin, 1987) is applied, implying that non-increasing numbers of records are checked by $k \geq 2$ subsequent auditors; the last of these is assumed to be infallible. Two sampling methods were discussed, called random and stratified sampling. In stratified sampling, previous classification results determine the next sample sizes for all classifications separately, while in random sampling they only determine the total sample size for the next auditor.

It was shown that both sampling methods lead to essentially the same MLE for the r population fractions p_{i_0}. However, if unstructural zeros occur, the MLE is not uniquely defined. Since unstructural zeros are much more likely to occur in case of random sampling, we advise stratified sampling for practical use. A further advantage is that the MLE in this case is unbiased.

For the specific case $r = k = 2$ with random sampling, our general formulae were shown to coincide essentially with the results of Tenenbein (1970) and Moors (1999). A new solution to the unstructural zeros problem was proposed having two advantages: it leads to a MLE with a smaller bias, and encompasses the solutions for the reduced models, where only one error type can occur.

In case error sizes, or relative error sizes (taintings) are observed instead of just error rates, continuous data are obtained. The special case of normally distributed observations with k subsequent auditors is being studied at the moment. Note that a distribution-free solution can be derived from the present paper by discretization of the continuous variable into r categories.

8 References

Bishop, Y. M. M., Fienberg, S.E., and Holland, P.W. (1975), *Discrete multivariate analysis: theory and practice*, Cambridge, MA:MIT Press.

Lehmann, E.L. and Casella, G. (1998), *Theory of point estimation*, Springer-Verlag, New York.

Little, R.J.A. and Rubin, D.B. (1987), *Statistical analysis with missing data*, John Wiley, New York.

Moors, J.J.A. (1999), *Double checking for two error types*, CentER Discussion Paper 9923, Tilburg University

Moors, J.J.A., van der Genugten, B.B. and L.W.G. Strijbosch (2000), Repeated audit controls, *Statistica Neerlandica* **54**, 3-13

Raats, V.M. and J.J.A. Moors (2000), *Double checking for two error types*, CentER Discussion Paper 2000120 KUB

Schafer, J.L. (1997), *Analysis of incomplete multivariate data*, Chapmann & Hall, London.

Tenenbein, A. (1970), A double sampling scheme for estimating from binomial data with misclassifications, *Journal of the American Statistical Association* **65**, 1350-1361

Tenenbein, A. (1971), A double sampling scheme for estimating from binomial data with misclassifications; sample size determination, *Biometrics* **27**, 935-944

Tenenbein, A. (1972), A double sampling scheme for estimating from inspection, *Technometrics* **14**, 187–202

9 Appendix A: Limit distribution

Write
$$\begin{cases} \Sigma^{(1)} = Cov(A^{(1)})/n_1 \\ \Sigma_c^{(j|j-1)} = Cov(A_c^{(j|j-1)})/n_c^{(j)} \end{cases}$$

with elements $f, g = 1, ..., r$:
$$\Sigma^{(1)}(f,g) = \begin{cases} \pi_{r-f}(1-\pi_{r-f}) & \text{if } f = g \\ -\pi_{r-f}\pi_{r-g} & \text{if } f \neq g \end{cases} \quad \text{and}$$
$$\Sigma_c^{(j|j-1)}(f,g) = \begin{cases} \pi_{r-f|c}(1-\pi_{r-f|c}) & \text{if } f = g \\ -\pi_{r-f|c}\pi_{r-g|c} & \text{if } f \neq g. \end{cases},$$

Then the asymptotic distributions of the MLE (4) are straightforward:
$$\sqrt{n_1} vec(\widehat{\Pi}^{(1)} - \Pi^{(1)}) \quad \xrightarrow{\mathcal{L}} \quad N(0, \Sigma^{(1)}),$$
$$\sqrt{n_c^{(j)}} vec(\widehat{\Pi}_c^{(j|j-1)} - \Pi_c^{(j|j-1)}) \quad \xrightarrow{\mathcal{L}} \quad N(0, \Sigma_c^{(j|j-1)})$$

If $\dfrac{N_c^{(j)}}{n_1} \xrightarrow{p} b_c^{(j)}$ with $b_c^{(j)}$ a constant depending on c,
$$\sqrt{n_1} vec(\widehat{\Pi}_c^{(j|j-1)} - \Pi_c^{(j|j-1)}) \quad \xrightarrow{\mathcal{L}} \quad N(0, \frac{\Sigma_c^{(j|j-1)}}{b_c^{(j)}}).$$

Since $\widehat{\Pi}^{(1)}$ and $\widehat{\Pi}_c^{(j|j-1)}$ are independent, they have an asymptotic multivariate normal distribution with a block-diagonal covariancematrix. The MLE for $\pi_{i_1 \dots i_k}$ is a function of the preceding estimators (1b): $\widehat{\Pi}_{i_1 \dots i_k} = \widehat{\Pi}_{i_1} \cdot \widehat{\Pi}_{i_2|i_1} \cdot \dots \cdot \widehat{\Pi}_{i_k|i_1 \dots i_{k-1}}$. Application of the deltamethod (see Casella and Lehmann (1999)) results in the asymptotic distribution of $\widehat{\Pi}_{i_1 \dots i_k}$. Relation (1(c)) and applying the deltamethod once more result in the asymptotic distribution of $\widehat{P}_{i_0}$ (5).

We only illustrate the whole procedure for the special case $r = k = 2$:

$$\sqrt{n_1} \begin{bmatrix} \widehat{\Pi}_1\text{-}\pi_1 \\ \widehat{\Pi}_0\text{-}\pi_0 \\ \widehat{\Pi}_{1|1}\text{-}\pi_{1|1} \\ \widehat{\Pi}_{0|1}\text{-}\pi_{0|1} \\ \widehat{\Pi}_{1|0}\text{-}\pi_{1|0} \\ \widehat{\Pi}_{0|0}\text{-}\pi_{0|0} \end{bmatrix} \quad \xrightarrow{\mathcal{L}} \quad N(0, \Sigma), \text{ with } \Sigma =$$

$$\begin{bmatrix} \pi_1\pi_0 & -\pi_1\pi_0 & 0 & 0 & 0 & 0 \\ -\pi_1\pi_0 & \pi_1\pi_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \pi_{1|1}\pi_{0|1}/b_1^{(2)} & -\pi_{1|1}\pi_{0|1}/b_1^{(2)} & 0 & 0 \\ 0 & 0 & -\pi_{1|1}\pi_{0|1}/b_1^{(2)} & \pi_{1|1}\pi_{0|1}/b_1^{(2)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{1|0}\pi_{0|0}/b_0^{(2)} & -\pi_{1|0}\pi_{0|0}/b_0^{(2)} \\ 0 & 0 & 0 & 0 & -\pi_{1|0}\pi_{0|0}/b_0^{(2)} & \pi_{1|0}\pi_{0|0}/b_0^{(2)} \end{bmatrix}$$

The deltamethod applied to relation (1(b)) results in

$$\sqrt{n_1}\begin{bmatrix} \widehat{\Pi}_{11}\text{-}\pi_{11} \\ \widehat{\Pi}_{10}\text{-}\pi_{10} \\ \widehat{\Pi}_{01}\text{-}\pi_{01} \\ \widehat{\Pi}_{00}\text{-}\pi_{00} \end{bmatrix} \quad \xrightarrow{\mathcal{L}} \quad N(0,\Delta), \text{ with } \Delta = B'\Sigma B,$$

$$\text{where } B = \begin{bmatrix} \pi_{1|1} & \pi_{0|1} & 0 & 0 \\ 0 & 0 & \pi_{1|0} & \pi_{0|0} \\ \pi_1 & 0 & 0 & 0 \\ 0 & \pi_1 & 0 & 0 \\ 0 & 0 & \pi_0 & 0 \\ 0 & 0 & 0 & \pi_0 \end{bmatrix}. \text{ So, } \Delta =$$

$$\begin{bmatrix} \pi_{1|1}\pi_{11}\pi_0 + \dfrac{\pi_{11}\pi_{10}}{b_1^{(2)}} & \pi_{11}\pi_{0|1}\pi_0 - \dfrac{\pi_{11}\pi_{10}}{b_1^{(2)}} & -\pi_{11}\pi_{01} & -\pi_{11}\pi_{00} \\ \pi_{11}\pi_{0|1}\pi_0 - \dfrac{\pi_{11}\pi_{10}}{b_1^{(2)}} & \pi_{0|1}\pi_{10}\pi_0 + \dfrac{\pi_{11}\pi_{10}}{b_1^{(2)}} & -\pi_{10}\pi_{01} & -\pi_{10}\pi_{00} \\ -\pi_{11}\pi_{01} & -\pi_{10}\pi_{01} & \pi_{1|0}\pi_{01}\pi_1 + \dfrac{\pi_{01}\pi_{00}}{b_0^{(2)}} & \pi_{0|0}\pi_{01}\pi_1 - \dfrac{\pi_{01}\pi_{00}}{b_0^{(2)}} \\ -\pi_{11}\pi_{00} & -\pi_{10}\pi_{00} & \pi_{0|0}\pi_{01}\pi_1 - \dfrac{\pi_{01}\pi_{00}}{b_0^{(2)}} & \pi_{0|0}\pi_{00}\pi_1 + \dfrac{\pi_{01}\pi_{00}}{b_0^{(2)}} \end{bmatrix}$$

Applying the deltamethod once again but this time to relation (1(c)) leads to the asymptotic distribution of $\widehat{P}_{i_0}$:

$$\sqrt{n_1}\begin{bmatrix} \widehat{P}_1\text{-}p_1 \\ \widehat{P}_0\text{-}p_0 \end{bmatrix} \quad \xrightarrow{\mathcal{L}} \quad N(0, B'\Delta B) \text{ where } B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Appendix B: Bias

Derivation of the bias of the modified estimator (12):

$$
\begin{aligned}
E(\widehat{P}_0^*) = \quad & E(E(\widehat{P}_0^*|N_0^{(2)})) \\
= \quad & Pr(N_0^{(2)}=0)E(\frac{A_0}{n_1}+\frac{A_1}{n_1}\frac{A_{10}}{N_1^{(2)}}|N_0^{(2)}=0)+ \\
& \sum_{n_0^{(2)}=1}^{n_2-1} Pr(N_0^{(2)}=n_0^{(2)})E(\frac{A_0}{n_1}\frac{A_{00}}{N_0^{(2)}}+\frac{A_1}{n_1}\frac{A_{10}}{N_1^{(2)}}|N_0^{(2)}=n_0^{(2)})+ \\
& Pr(N_0^{(2)}=n_2)E(\frac{A_0}{n_1}\frac{A_{00}}{N_0^{(2)}}|N_0^{(2)}=n_2) \\
= \quad & Pr(N_0^{(2)}=0)(\frac{(n_1-n_2)\pi_0}{n_1}+\frac{(n_1-n_2)\pi_1+n_2}{n_1}\pi_{0|1})+ \\
& \sum_{n_0^{(2)}=1}^{n_2-1} Pr(N_0^{(2)}=n_0^{(2)})(\frac{(n_1-n_2)\pi_0+n_0^{(2)}}{n_1}\pi_{0|0}+\frac{(n_1-n_2)\pi_1+n_1^{(2)}}{n_1}\pi_{0|1})+ \\
& Pr(N_0^{(2)}=n_2)(\frac{(n_1-n_2)\pi_0+n_2}{n_1}\pi_{0|0}) \\
= \quad & Pr(N_0^{(2)}=0)(1-\frac{n_2}{n_1})\pi_{01}-Pr(N_0^{(2)}=n_2)(1-\frac{n_2}{n_1})\pi_{10}+ \\
& E(\frac{(n_1-n_2)\pi_0+N_0^{(2)}}{n_1}\pi_{0|0}+\frac{(n_1-n_2)\pi_1+N_1^{(2)}}{n_1}\pi_{0|1}) \\
= \quad & p_0+(1-\frac{n_2}{n_1})(\pi_1^{n_2}\pi_{01}-\pi_0^{n_2}\pi_{10})
\end{aligned}
$$

The bias of $\widehat{P}_0$ (11) can be derived in a similar way.